COMPARATIVE EVALUATION OF A STOCHASTIC PARSER ON SEMANTIC AND SYNTACTIC-SEMANTIC LABELS

Wolfgang Minker

University of Ulm
Dept. of Information Technology
89081 Ulm/Donau, Germany
wolfgang.minker@e-technik.uni-ulm.de

Abstract

This paper deals with the evaluation of a stochastic component for natural language understanding alternatively trained on semantic and syntactic-semantic labels. The parser uses semantically-labeled speech data gathered using the LIMSI-ARISE interactive speech system for train travel information retrieval in French. The study shows that introducing additional and coherent information into the semantic corpus allows to relatively improve the semantic frame accuracy of the parser by up to 16.5%. The more complex models yielding a high number of parameters are justified, as long as they convey significant information.

1. Introduction

The European project ARISE (Automatic Railway Information Systems for Europe) supports the development of train schedule inquiry services by telephone (Blasband, 1998). A spoken language system that operates for this task automatically deals with over 200 million routine inquiries that are routed every year to the European railway information centers. Around 20% of these inquiries remain without response as there are not enough human operators. ARISE phone servers have been developed for the Dutch, Italian and French railway operators.

In the framework of this project, LIMSI-CNRS has improved its interactive speech system for train travel information retrieval in French, initially designed for the projects MASK and RAILTEL (Lamel *et al.*, 1995). The system employs a rule-based parsing method for the semantic analysis. The phone server ARISE provides information about train schedules, services and fares between 600 French and European cities (Lamel *et al.*, 1998). During the data collection campaigns, that provide spoken language corpora for system development, potential users have been recorded using a prototype system.

A stochastic component for natural language understanding, initially developed as a part of a spoken language system for the information retrieval applications ATIS and MASK (Minker, 1998), has been ported to the French ARISE task (Minker, 1999b). The stochastic models used by the parsing component are derived from the automatic analyses of large corpora of naturally uttered sentences along with their semantic representations. Stochastic methods for natural language understanding have also been applied in the BBN-HUM (Schwartz *et al.*, 1996) and the AT&T-CHRONUS (Levin & Pieraccini, 1995) systems.

In ATIS and MASK, the stochastic parsing component was trained and evaluated on a purely semantic case grammar representation (Minker, 1998). Based on a structure spotting, such a semantic grammar is robust and well adapted to spontaneous human-machine interaction. However, the robustness is likely to turn into a drawback, if the semantic analysis ignores information that is propagated by syntactic relations. Introducing additional syntax information, whose complexity is well adapted to the size of the stochastic model, may disambiguate and therefore improve the decoding.

In the work described in this paper, the stochastically-based parsing component takes advantage of the availability of the speech data gathered using the ARISE end-to-end system. We study the impact on the frame accuracy when introducing syntax information into the stochastic case grammar parser. Limited to isolated utterance transcriptions the model does not account for speech recognizer output lattices nor the dialog context.

In the following section, we briefly introduce the semantic case grammar formalism. Section 3 describes the stochastic parsing method and the introduction of syntax information. Evaluation results are discussed in Section 4.

2. Semantic Case Grammar

In the domain of spoken language parsing, failures occur due to false starts, repetitions and ill-formed utterances. A well-known approach to extract semantic information from spontaneous speech is based on *case frames* (Fillmore, 1968; Bruce, 1975). The concept of the frame is identified by one or several reference words in the sentence. The attributes of the frame (cases) are instantiated using specific markers that represent local syntactic rules. An example of a semantic frame for the ARISE application is given in Figure 1. Utterance meaning is described using a set of attribute-value pairs (Minker, 1999). Several levels of attributes may be distinguished:

- **Arguments:** They allow a description of the query type uttered by the user. Any low-level attribute of the semantic representation (cf. below) may be used as a value of the argument. In the example, the value is fare, instantiated by *prix* (*fare*).

- **Dialog-related attributes:** They capture introducing, closing, politeness and connective forms and may also contain the response of the user to a proposition made by the system. In the example, (+/response) and (+/dialog) are dialog-related attributes.

- **Task-related attributes:** They contain elementary information, such as dates, times, localities, etc. In the example, (?/class) is a task-related attribute.

Utterance:
j'aimerais connaître le prix en deuxième classe s'il vous plaît
(*I would like to know the fares second class please*)

Semantic frame:
(?/argument:fare)
(?/class): *deuxième*
(+/dialog): *s'il vous plaît*

Figure 1: Example of an utterance with the corresponding semantic frame in ARISE.

A **mode** is associated to each attribute of the frame. Informative (+/), it indicates whether the user provides information, negative (-/), it indicates an auto-correction of the associated attribute and interrogative (?/), it indicates that the user asks for information. In the example of Figure 1, the mode is interrogative for (?/argument) and (?/class), and informative for (+/dialog). Changing modes within the frame is equivalent to breaking it down into sub-frames. These represent a certain rhetoric unit (order of the sentence).

3. Stochastic Parsing Method

The semantic analyzer described in the following makes use of stochastic modeling techniques for learning and processing the case grammar introduced above. The mappings which link the semantic representation with words from the input stream are referred to as *semantic* or, alternatively, *syntactic-semantic labels*.

Processing steps Two main processing steps can be described in the stochastic parsing component. During training, the parameter estimator establishes the stochastic model from a large number of transcribed utterances and the corresponding label sequences. In the decoding or testing mode the semantic decoder, implemented as a Hidden Markov Model - HMM (Rabiner & Juang, 1986) outputs the most likely label sequence given a test utterance transcription.

Stochastic modelling Relative occurrences of model states and observations are used to establish the HMM. The labels are defined as the states s_j. All states, such as (+/number-ticket), (+/null) and (+/command) may follow each other; thus the model is ergodic.

Following the HMM theory, semantic decoding consists of maximizing the conditional probability $P(S|O)$ of some state sequence S given the observation sequence O. The pre-processed words in the utterances are defined as the observations o_m.

Based on the model, the most likely state sequence is determined using the *Viterbi algorithm* (Rabiner & Juang, 1986). Given a significant number of model parameters, a *back off* technique (Katz, 1987) allows an adequate probability estimation of rare observation and state occurrences. In this work, the back off models were calculated using the CMU-toolkit (Rosenfeld, 1995).

Knowledge representation The stochastic method deals with word and label sequences. Therefore the frame representation (Figure 2(a)) needs to be aligned with the input utterance (Figure 2(b)). Markers, mode identifiers (such as *connaître* (*know*) ↦ (mode:?)) in Figure 1 and irrelevant labels that are not explicit in the frame are therefore represented here.

je souhaiterais réserver
une place de Paris à Bordeaux

(+/command): *réserver*
(+/number-ticket): *1*
(+/place-from): *Paris*
(+/place-to): *Bordeaux*

(a)

(I)	(II)		(III)
je	⇐ (+/null)		[personal-pronoun]
souhaiterais	⇐ (+/null)		[verb]
réserver	(+/command)		[verb]
une	(+/number-ticket)		[indefinite-article]
place	⇐ (+/m:number-ticket)	⇐	[noun]
de	⇐ (+/m:from)		[preposition]
Paris	(+/place-from)		[name]
à	⇐ (+/m:to)		[preposition]
Bordeaux	(+/place-to)		[name]

(b)

Figure 2: Semantic frame representation for *je souhaiterais réserver une place de Paris à Bordeaux* (*I would like to book one seat from Paris to Bordeaux*) (a); input utterance (b-I); attributes, markers and irrelevant labels to build a semantic sequence (b-II); sequence of syntactic labels (b-III). The sequences (II) and (III) need to be merged in order to obtain the syntactic-semantic labels.

The markers in sequence (II) help to disambiguate the decoding. In the example *une place*, the word *place* (*place*) ↦ (+/m:number-ticket) designates *une* (*one*) to be a (+/number-ticket). A total of 13 semantic markers has been used. Also in sequence (II), the irrelevant label (+/null) corresponds to words that do not yield any particular semantic function in the context of the utterance, e.g., *je souhaiterais* (*I would like*).

The sequence (III) in Figure 2(b) shows the syntactic annotation of the utterance. It has been performed by SYLEX, a syntactic analyzer for the French language (Constant, 1991). Based on the identification of syntactic groups or chunks, SYLEX labels each individual word of the input sentence with one of the 25 syntactic categories.

The syntactic-semantic labels are obtained after merging the sequences (II) and (III) in Figure 2(b). Each label represents the semantic function of the word along with its syntactic role in the utterance.

Figure 3 illustrates how the syntactic-semantic labels are used in the model.

Stochastic methods require substantial amounts of data for the estimation of their parameters. Corpora of spoken language are still limited in size, a fact that is problematic because events rarely observed in the training data are not adequately modeled. As a result, the estimates may become

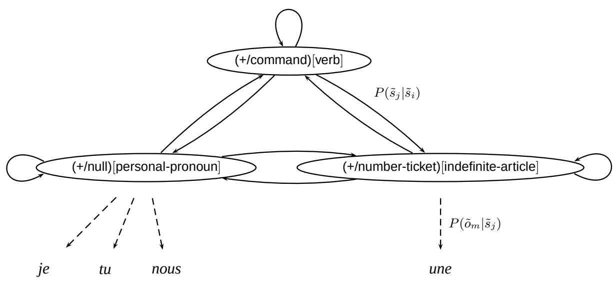
(+/command)[verb]
$P(\tilde{s}_j|\tilde{s}_i)$
(+/null)[personal-pronoun]
(+/number-ticket)[indefinite-article]
$P(\tilde{o}_m|\tilde{s}_j)$
je
tu
nous
une

Figure 3: Syntactic-semantic labels in the HMM.

unreliable. Therefore the data sparseness requires matching the model size to the amount of training data available using utterance pre-processing strategies. The lexical analysis and category unification are based on lookup tables, established after a manual analysis of the training corpus.

Lexical analysis: Inflected words are replaced with their corresponding base forms and semantically-related words are clustered. E.g., *accepte*, (*accept*), *d'accord* (*alright*) are attributed to the word cluster $\{$*accepter*$\}$, and *type* (*type*), *types* (*types*) are replaced by $\{$*type*$\}$. Non-relevant or out-of-domain words (with respect to the given application) are assigned to a $\{$*filler*$\}$ category and removed from the corpora as they do not contain nor propagate any significant information.

Category unification: In the domain of train timetable database information retrieval, a significant number of lexical entries correspond to database values, which can also sometimes be clustered. Typical word categories in the train travel domain deal with dates, times and localities, e.g.:

> *euh je souhaite un départ de* ***Redon*** *vers* ***Montpellier*** *pour le* ***vingt-cinq janvier*** *à* ***dix-neuf*** *heures*
> (*I would like to leave from* ***Redon*** *to* ***Montpellier*** *on* ***January 25th*** *at* ***7 o'clock pm***)

11 task-related categories */COMFORT/, /STATION/, /RELATIVE-DAY/, /WEEKDAY/, /HOLIDAY/, /MONTH/, /NUMBER/, /ORDINAL/, /TIME-SLOT/, /SERVICE/* and */TRAIN-TYPE/* have been defined. Using lexical simplification, the above example is preprocessed to:

> $\{$*filler*$\}$ $\{$*filler*$\}$ $\{$*filler*$\}$ $\{$*départ*$\}$ *de /STATION/ vers /STATION/ pour le /ORDINAL/ /MONTH/ à /NUMBER/ heures*

Even though the lexical simplification reduces the quality, it also reduces the complexity of the stochastic model. After this pre-processing, the lexicon size decreases considerably from 1,859 to 348 words.

Training corpus In order to estimate the model parameters the stochastic parser requires a corpus of utterance transcriptions labeled with the corresponding semantic or syntactic-semantic categories. A common semi-automatic procedure has been used to label the utterances (Minker, 1998).

The stochastic model has been trained using transcriptions of 14,500 utterances along with their sequences of semantic and, alternatively, syntactic-semantic labels. Table 1 provides an overview of the characteristics of both training corpora.

	Lexicon size		*#labels*	
#utts	*raw*	*pre-processed*	*semantic*	*syntactic-semantic*
14,500	1,859	348	256	694

Table 1: Data characteristics of the ARISE training corpora.

We note the significant decrease in the lexicon size after the utterance pre-processing. Notably the category */STATION/* replaces more than 650 station names (35% of the lexicon) by a single variable.

The development of a stochastic component is the result of trying to optimally combine several countervening factors. These include the complexity and the quality of the stochastic model, the type of application, the type of syntactic and/or semantic representation and, finally, the amount of training data available for the parameter estimation. The major problem with a stochastic method is to find a good balance between the sparse data and the model complexity. An increase in the complexity is justified as long as the data amount is sufficient and the information conveyed by the topology improves the quality of the model.

This being said, the utterance pre-processing in ARISE reduces the complexity of the parameter model, but also its quality in terms of the information conveyed. In turn, the fact of replacing semantic by syntactic-semantic labels increases the model complexity, but also introduces valuable additional information.

Table 2 shows the number of observation probabilities, which is an indicator for the model size, as a function of the label type and the type of utterance pre-processing. The use of pre-processed utterances along with syntactic-semantic labels results in a lower number of parameters (186,829) compared to the use of purely semantic labels in combination with raw data (447,513).

	$\#P(o_m|s_j)$ occurring in	
Label type	*raw data*	*pre-processed data*
semantic	447,513	69,076
syntactic-semantic	1,131,699	186,829

Table 2: Number of observation probabilities $P(o_m|s_j)$ as a function of the label type and the type of utterance pre-processing employed.

4. Performance Assessment

The stochastic parsing component has been evaluated on the parsing output of 500 utterance transcriptions that have not been used throughout the component training. The reference representations have been obtained after an initial processing of the parser on the transcribed speech data followed by a manual correction.

The state sequences used for training (and therefore the parser output produced in the test) contain alternatively semantic and syntactic-semantic labels. However, the error rates were measured on the semantic frame level (Table 3), that only accounts for labels which are significant for further processing (c.f. the frame in Figure 1(a)). Consequently, the markers and irrelevant labels, but also the syntactic labels have been removed from the sequences prior to the evaluation.

	Semantic frame error (%) after training on	
Label type	*raw data*	*pre-processed data*
semantic	30.6	25.4
syntactic-semantic	26.4	21.2

Table 3: Performance evaluation results of the stochastic parser on the semantic frame level on raw and pre-processed utterance transcriptions. The model was trained and evaluated alternatively on semantic and syntactic-semantic labels.

In both set-ups, i.e. training and testing the component alternatively on raw and pre-processed utterances, the introduction of additional syntax information allows to obtain a relative performance gain of respectively 13.7% (from 30.6% down to 26.4%) and 16.5% (from 25.4% down to 21.2%).

5. Summary and Conclusions

We have described the impact of introducing syntax information in a stochastic parsing component that makes use of a semantic case grammar formalism. The parser operates in the French ARISE task, an application that supports the development of schedule inquiry services by telephone. The semantic analyzer makes use of stochastic modeling techniques for implementing the case grammar. The mappings which link the semantic representation with words from the input stream are referred to as semantic or, alternatively, syntactic-semantic labels. The semantic annotation of the corpus was performed using an iterative labeling approach. The syntactic annotation of the corpus has been obtained using SYLEX, a syntactic analyzer for the French language. The syntactic-semantic labels have been obtained by merging individual semantic and syntactic label sequences.

The stochastic parser has been evaluated on setups with both semantic and syntactic-semantic label types. Introducing syntax information increased the relative performance by up to 16.5%. We conclude from these experiments, that the fact of introducing additional and coherent information into the semantic corpus allows to improve the performance of the parser. Complex models yielding a high number of parameters are justified, as long as they convey significant information.

Certain aspects of the presented method could be further investigated and expanded. The procedures used to increase performance may be validated on other corpora and less restricted tasks. Furthermore, the initial use of statistical modeling for the semantic analysis was not integrated with the speech recognition, dialog and response generation components of a spoken language system. The results are presented on transcribed sequences only. A first step in this direction would be to compare the component performance using utterance transcriptions with those obtained when using real speech recognizer output either for component training and testing or testing only.

6. Acknowledgements

The research has been carried out during a stay at the Spoken Language Processing Group at LIMSI-CNRS. The author gratefully acknowledges the contribution of Sophie Rosset, Yannick de Kercadio, Aggeliki Fotopoulou and Gabriel Illouz who provided assistance in the experiments reported in this article.

7. References

Blasband M. (1998), “Speech Recognition in Practice: The ARISE Project,” In: La Lettre de l’IA.

Bruce B. (1975), “Case Systems for Natural Language,” Artificial Intelligence, Vol. 6.

Constant, P. (1991), “Analyse Syntaxique par Couches,” Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris.

Fillmore, Ch. J. (1968), “The he case for case,” In: Universals in Linguistic Theory, Bach Emmon and Harms Robert T. (eds.), Holt and Rinehart and Winston Inc.

Katz, S. M. (1987), “Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer,” IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 35(3).

Lamel L.F., Rosset S., Gauvain J.L., Bennacef S.K., Garnier-Rizet M. and Prouts B. (1998), “The LIMSI ARISE System,” Interactive Voice Technology for Telecommunications Applications, IVTTA.

Lamel L.F., Bennacef S.K., Bonneau-Maynard H., Rosset S. and Gauvain J.L. (1995), “Recent Developments in Spoken Language Systems for Information Retrieval,” ESCA Workshop on Spoken Dialogue Systems.

Levin E. and Pieraccini R. (1995), “CHRONUS - The Next Generation,” ARPA Workshop on Spoken Language Technology.

Minker W. (1998), “Stochastic versus Rule-based Speech Understanding for Information Retrieval,” Speech Communication, Vol. 25(4).

Minker W. (1999), “Compréhension de la parole spontanée,” L’Harmattan.

Minker W. (1999b), “Stochastically-based semantic analysis for ARISE - Automatic Railway Information Systems for Europe,” Grammars, Vol. 2(2).

Rabiner L.R. and Juang B.H. (1986), “An introduction to Hidden Markov Models,” IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 3(1).

Rosenfeld, R. (1995), “The CMU Statistical Language Modeling Toolkit, and its use in the 1994 ARPA CSR Evaluation,” ARPA Workshop on Spoken Language Technology.

Schwartz R., Miller S., Stallard D. and Makhoul J. (1996), “Language Understanding Using Hidden Understanding Models,” International Conference of Speech and Language Processing, ICSLP.